我国数据新闻研究领域近十年的动态演进
——基于 CiteSpace 的数据挖掘分析

胡岩岩
（新华报业传媒集团，江苏 南京 210005）

摘要：【目的】数据新闻作为一种跨领域、多学科的新闻报道方式，近年来受到广泛关注，因此很有必要梳理其动态演化路径和发展趋势，为我国数据新闻领域前沿性问题的探索提供一定基础。【方法】文章以中国学术期刊网络出版总库(CNKI)收录的数据新闻相关文献为研究对象，运用计量软件 CiteSpace 绘制出知识图谱进行可视化分析。【结果】旨在对 2013 年以来近十年国内数据新闻研究的发展态势、研究主题和热点变化进行挖掘分析。【结论】研究发现，在经过 2013—2016 年的短期井喷式快速发展之后，从 2017 年开始，相关研究量进入回落阶段，研究逐渐趋于理性。相关图谱显示，近十年来国内对数据新闻的研究热点主要聚焦于数据新闻的概念、生产流程、价值影响以及人才培养等方面，并且逐渐从早期的对基本理念、可视化实践等的研究向数据新闻的叙事性回归、人才培养以及新技术应用发展等研究转变。

关键词：数据新闻；知识图谱；CiteSpace；叙事性回归；人才培养　　**中图分类号：**G240　　**文献标识码：**A
文章编号：1671-0134（2023）04-053-05　　**DOI：**10.19483/j.cnki.11-4653/n.2023.04.011
本文著录格式：胡岩岩 . 我国数据新闻研究领域近十年的动态演进——基于 CiteSpace 的数据挖掘分析 [J]. 中国传媒科技，2023（04）：53-57.

数据新闻是在云计算、大数据、人工智能、移动互联网等新技术快速发展的背景下应运而生，作为一种全新的跨领域、多学科的新闻报道方式，一经出现，便受到广泛的关注。2009 年，英国《卫报》在全球第一个设立数字新闻部。2010 年，在阿姆斯特丹举行的首届“国际数据新闻”圆桌会议，从制作流程的角度对数据新闻做了详尽的阐释。2011 年，始于伦敦举行的 MozFest 大会 48 小时工作坊，多名数据新闻倡导者及专家共同编撰了第一本数据新闻制作指南《数据新闻手册》。2012 年全球数据新闻奖设立。2013 年《数字新闻手册》中译版问世。2013 年 3 月新华网成立数据新闻部的官方网站，并推出数据新闻专栏，是国内最早一批进行数据新闻实践的媒体。2013 年 10 月，财新网成立财新数据可视化实验室，打造 “数字说”频道。2015 年，首届中国数据新闻大赛举办。此外还有澎湃新闻的“美数说”，网易的“数读”栏目等，近十年来数据新闻无论在内容生产、叙事模式，还是知识生产等方面都在持续探索和创新尝试中不断发展。

基于数据新闻被学界和业界广泛关注的现状及在当今的重要意义，本文采用国际潮流的科学研究评价方法—文献计量法，以 CNKI 中数据新闻领域核心期刊相关文献为研究样本，借助信息可视化软件 CiteSpace，挖掘、分析、梳理和绘制知识图谱来深度了解中国数据新闻领域的研究现状，明确数据新闻研究的动态演化路径和发展趋势，为中国数据新闻领域前沿性问题的探索提供一定基础。

1. 研究设计

1.1 研究方法

CiteSpace 是由陈超美教授研发的一款用于辅助文献计量分析的可视化软件 [1]，它能够将既定研究领域内发展进程中重要的信息变化通过可视化的手段展现出来，呈现出该领域的关键文献、研究热点、发展趋势等信息。本研究通过使用 CiteSpace 的词频共现分析、聚类分析、研究前沿探测等基本功能，呈现出数据新闻领域的研究热点和发展趋势，以期对近十年来我国数据新闻研究的动态演进有详尽的了解。

1.2 数据来源

本文数据来源于 CNKI，检索方式为高级检索，检索词选择主题“数据新闻”或者主题“数据驱动新闻”，发文时间限定为 2013—2022 年，期刊来源类别框定为“北大核心”“CSSCI”和“CSCD”，学科设定为“新闻与传媒”。检索后，筛除无效文献，最终获得期刊文献 751 篇。将文献数据导出，转换及保存后进行计

量分析。

2. 我国数据新闻研究现状分析

2.1 发文量情况分析

从近十年来中国知网核心期刊有关数据新闻研究的年度发文量维度出发，绘制成如图 1 所示的发文量趋势图。

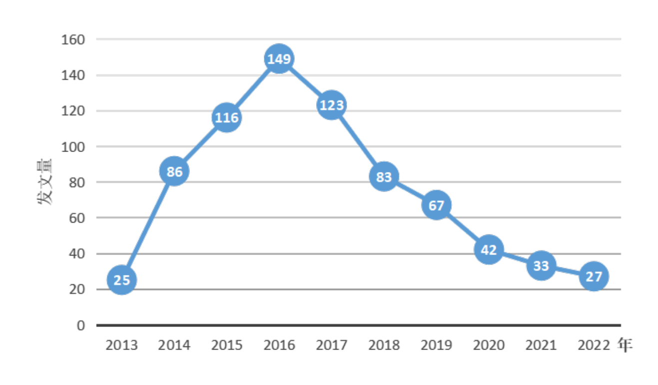


图 1 2013—2022 年我国数据新闻文献数量变化

从图 1 可以看出，有关“数据新闻”研究的文献发表数量，从 2013 年开始进入快速增长的阶段。也正是在 2013 年，诞生于 2011 年伦敦 Mozilla Festival 的 48 小时工作坊的全球第一本数据新闻制作指南《数字新闻手册》中译版问世，从此数据新闻开始在中国流行起来。从 2013 年到 2016 年年发文量由 25 篇飙升至峰值 149 篇，相关研究在 2013 年之后便呈现出井喷式的短期暴增趋势，这一阶段国内研究学者对数据新闻的研究热度日益高涨，研究内容不断深入，涉及的主题也在不断拓宽。而从 2017 年开始，国内数据新闻研究进入回落阶段，核心期刊发文量下滑，由 2017 年的 123 篇，降至 2022 年的 27 篇，国内数据新闻研究进入“降温”回落阶段，研究逐渐趋于理性。

2.2 发文作者分析

研究作者的发文量能够反映出该领域学术资源的分布情况。根据 CiteSpace 统计分析结果，绘制出 2013—2022 年数据新闻主要发文作者及发文量（表 1）。研究发现，发文量较高的作者多为来自各大高校的新闻传播学界人士。

作者间的合作方面，本研究通过 CiteSpace 工具绘制 2013-2022 年间数据新闻发文作者分布的知识图谱（图 2），其中节点之间的连线代表合作关系，发文量较高的作者显示为较大节点。分析图 2 可知，国内数据新闻研究论文的作者合作网络均比较松散，更倾向于独立产出成果，该领域尚未形成显著的群体效应。不过在这种更加开放自由的研究模式下，更易于积累多样化的研究成果，从而实现多元创新。

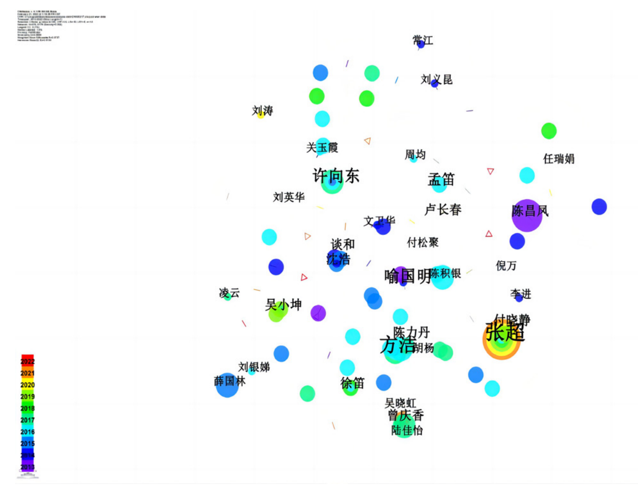


图 2 2013—2022 年我国数据新闻研究作者合作图谱

3. 我国数据新闻研究热点及前沿分析

表 1 2013-2022 年数据新闻主要发文作者及发文量

序号	作者	发文篇数	首次发文时间（年）	作者单位
1	张超	14	2016	山东大学（威海）文化传播学院
2	方洁	11	2013	中国人民大学新闻学院
3	喻国明	7	2013	中国人民大学新闻学院
4	许向东	7	2015	中国人民大学新闻学院
5	孟笛	5	2016	华东师范大学传播学院
6	卢长春	4	2017	曲阜师范大学传媒学院
7	曾庆香	4	2017	中国传媒大学新闻学院
8	付晓静	4	2015	武汉体育学院新闻传播学院
9	吴小坤	4	2017	华南理工大学新闻与传播学院
10	陈力丹	4	2016	四川外国语大学新闻传播学院
11	徐笛	4	2016	复旦大学新闻学院
12	谈和	4	2014	中国传媒大学传播研究院
13	陈昌凤	4	2013	清华大学新闻与传播学院
14	沈浩	4	2014	中国传媒大学新闻学院

3.1 关键词共现分析

关键词是对一篇文献主要内容及核心主题的高度概括，如果某个关键词在该研究领域的文献中多次出现，则该关键词常被用来反映热点研究内容。共现分析图谱即是用一个标有关键词的圆圈作为节点，节点之间的连线表示关键词之间存在共现关系，圆圈的半径代表该关键词的出现频率。[2] 将 751 篇文献数据导入 CiteSpace 软件运行之后得到如图 3 所示的关键词共现图谱。图谱中可见关键词共现网络由 402 个关键词节点和 475 个连线组成。进一步研读图 4 后可以发现，近十年来，数据新闻的研究热点主要集中在数据新闻理念，以及大数据、可视化、媒体融合、新闻生产与传播、数据素养、人才培养等方面。

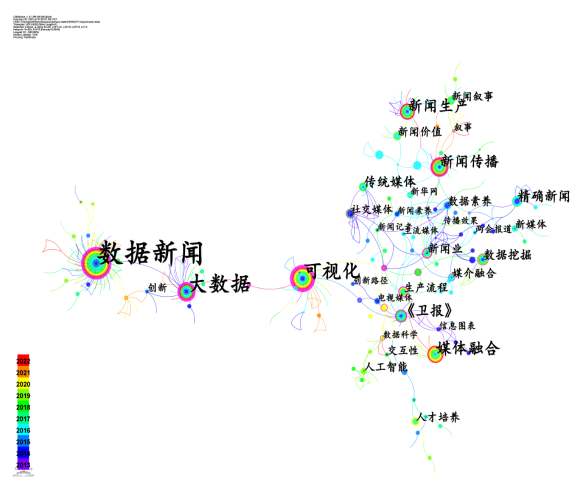


图 3　2013-2022 年我国数据新闻研究关键词共现图谱

3.2 研究知识聚类

为了进一步归纳关键词之间的联系、总结数据新闻领域的研究主题，本研究在关键词共现分析的基础上对文献中关系紧密的关键词进行聚类分析，所形成的聚类簇代表当前热点研究领域。聚类簇采用聚类模块度（Modularity Q）和平均轮廓值（Mean Sihouette）判断可信度。本研究聚类分析结果如图 4 所示，对国内数据新闻研究关键词聚类的模块度为 0.8609，平均轮廓系数 S 为 0.9727，说明聚类效果较好，聚类图谱合理。2013—2022 年我国数据新闻相关的研究共得到 17 个自动聚类。为厘清数据新闻研究的知识体系和发展脉络，本文结合图谱分析结果和文献精读情况，将 17 个自动聚类结果归纳为四大类研究热点。

第一类为数据新闻的基本理念及创新发展研究，主要包括聚类图中的“#0 数据新闻”“#3 大数据”。2013 年是我国的大数据元年，同时《数据新闻手册》中译版于同年正式引入中国，数据新闻由此拓展了新闻学基本理念。国内部分学者在梳理国际数据新闻研究的过程中，总结并提出了数据新闻的各项特征要素等基本理念。方洁、颜东于 2013 年发表论文讨论了数据新闻的内涵、特征，并理清了数据新闻的相关概念。[3] 任瑞娟提出，数据新闻的要素包括数据要素、超链接要素、交互化要素、场景化要素、视化要素和语义化要素。[4]

第二类为数据新闻的生产流程研究，这类研究主要包括聚类图中的“#2 可视化”“#6《卫报》”“#7 精确新闻”“#8 传统媒体”“#9 新闻叙述”“#10 数据科学”“#11 媒体数据”“#12 数据挖掘”“#15 新闻制作”“#16 开放性”。相关论文主要关注数据新闻的数据源、数据挖掘分析、叙事模式、数据呈现，以及数据新闻与新技术应用的研究。首先，海量的数据源是数据新闻的前提，通过数据将多种信息联系起来，增强新闻报道精确性和说服力。刘英华等认为开放数据是一类可以被任何人免费使用、再利用、再分发的数据。[5] 毕秋灵指出，开放数据主要由政府机构、高校等科研机构，以及非政府组织提供，具有公开性、可得性、完整性、即时性等特点。[6] 其次，技术裹挟着数据，数据建模、数据挖掘、NLP 语义分析、知识图谱、推荐算法等均是数据新闻应用的关键技术。[7] 同时这些新技术的应用，也展现了数据新闻的发展趋势。最后，数据新闻以可视化路径创新了新闻叙事模式，许向东认为数据新闻报道中“叙事者”部分物化为“数据”，叙事结构从二维到立体。可视化呈现对繁冗、巨量的数据信息进行勾陈、梳理、归纳，使其形成一个相对流畅的视觉叙事链提升阐述功能，同时能增强用户的交互体验。[8]

第三类为数据新闻的价值及影响研究，相关论文主要包括“#4 新闻传播”“#5 新闻价值”“#14 媒体融合”。数据新闻在智能媒体的时代背景下作为一种充满活力的创新新闻样态，将成为消息、通讯、特写、深度报道等传统新闻报道样态的良好补充。任瑞娟在《预测与发现——数据新闻理论与实践》提到大数据时代数据新闻的一个重要价值是具备预测与发现功能。[9] 通过对数据进行结构化、知识化处理，探索出有意义的数据联系，并基于现实数据分析，审视现在和过去，对事件的未来发展状况进行可能性预测，真正做到从“事后总结”到“事前策划”的新闻变革。数据新闻为主流媒体和新闻记者提供了新机遇，使其专业内容生产者的身份得以重新建构。[10]

第四类为数据新闻的人才培养和数据素养研究，这类研究主要在图中的“#1 数据素养”“#13 技术赋能”。相关论文主要关注数据新闻人才培养模式和新闻传播人才的数据素养等方面。许向东在《对中美数据新闻人才培养模式的比较和思考》一文通过分析中美数据新闻人才培养模式及其特点，探索适合我国新闻教育和传媒业实践特点的数据新闻人才培养模式。[11] 方洁、高璐在《数据新闻：一个亟待确立专业规范的领域——基于国内五个数据新闻栏目的定量研究》中通过对国内五个数据新闻栏目一年的报道内容分析研究发现，国内数据新闻领域亟待确立专业规范，我国数据新闻从业人员的数据素养亟待提高。[12]

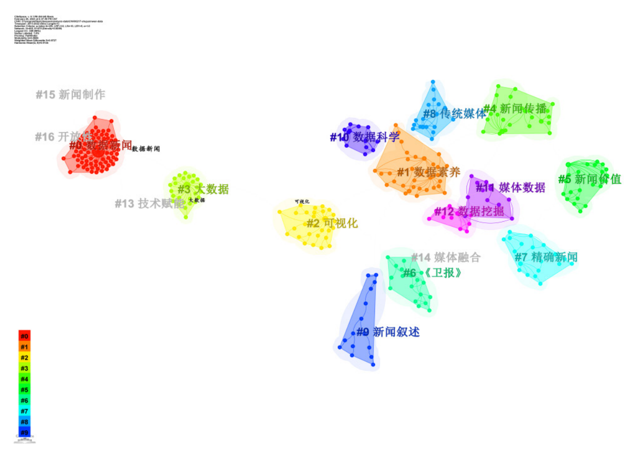


图 4　2013—2022 年我国数据新闻研究关键词聚类图谱

3.3　关键词时间线视图分析

分析关键词时间图可以明确国内数据新闻领域近十年的研究脉络。同一聚类的节点按照时间顺序被排布在同一水平线上，时间轴位于视图上方，从左向右顺序分布。聚类中包含的文献数量越多，代表该聚类在领域中越重要；时间跨度越长代表该聚类中的研究领域越早、持续性也越长。本文对关键词进行进一步分析，选择“Timeline View”进行时间线视图可视化，制作如图 5 所示的 2013—2022 年我国数据新闻研究关键词时序图谱。

图中可以看出，“# 数据新闻”“# 可视化”“# 大数据”“# 新闻价值”这几个聚类产生较早，并且自出现便一直贯穿于数据新闻的研究中。数据新闻的基本理念，具体的可视化生产实践，以及产生的传播效果和价值等多个早期产生的研究热点，已经形成了稳定的方向，成为学者研究数据新闻时长期关注的主题。从 2017 年开始，部分研究热度开始呈现下降趋势，重点也开始转移。主要研究放在了数据新闻的新闻叙事性，以及数据科学、技术赋能等方面，实现了从基本理念，大数据可视化实践等的研究到数据新闻的叙事性回归、人才培养等研究的转变。

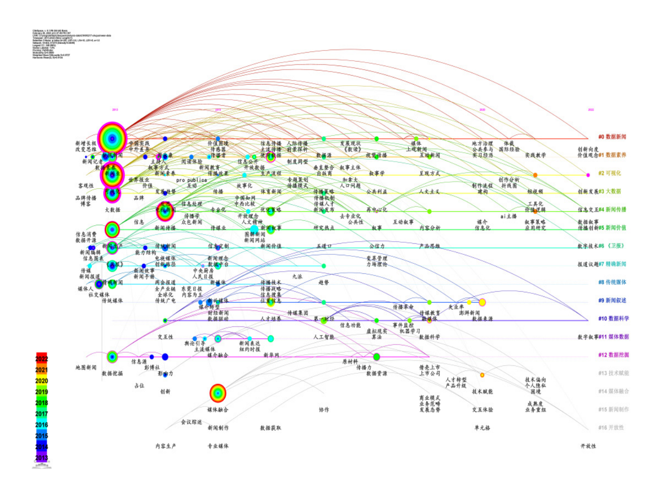


图 5　2013-2022 年我国数据新闻研究关键词时序图谱

3.4　研究前沿分析

突现词（Burst）是指在一定时间内频次突然呈现出显著增长的关键词。通过突变词可以了解研究的前沿情况、研究焦点的转变和最新的研究热点动态，并帮助预测该领域后续的发展趋势，探索其新的可能的前沿研究问题。本文利用 CiteSpace 软件的“Burstness”功能得到 2013—2022 年我国数据新闻研究关键词突现图谱，见图 6，其中 Year 代表该关键词的出现时间，Begin 和 End 分别代表该关键词成为研究热点的开始和结束。



Top 19 Keywords with the Strongest Citation Bursts

Keywords	Year	Strength	Begin	End	2013 - 2022
大数据	2013	4.6	2013	2014	
社交媒体	2013	2.75	2013	2014	
信息图表	2013	2.4	2013	2014	
新闻编辑	2013	2.06	2013	2014	
数据素养	2013	1.66	2013	2016	
数据库	2013	1.55	2013	2014	
电视媒体	2014	2.15	2014	2015	
创新	2014	1.62	2014	2015	
创新路径	2014	1.45	2014	2015	
新媒体	2015	1.71	2015	2017	
新闻教育	2015	1.46	2015	2016	
新华网	2016	2.14	2016	2017	
新闻价值	2016	1.82	2016	2017	
人工智能	2017	3.83	2018	2019	
叙事	2018	2.13	2018	2022	
数据科学	2019	2.31	2019	2022	
互动新闻	2019	1.65	2019	2020	
人才培养	2016	1.48	2019	2020	
技术赋能	2020	2.36	2020	2022	

图 6　2013—2022 年我国数据新闻研究关键词突现图谱

从时间上来说，最早出现的突现词为“大数据”“社

交媒体”“信息图表”“新闻编辑”“数据库”等。相关研究主要体现在数据新闻本身理念及产生的影响，以及关联的大数据、可视化实践等方面。近几年来，尤其是从2018年开始，产生了“人工智能”“叙事”“数据科学”“技术赋能”等新的突变词，说明人工智能、数据科学等在数据新闻中的应用发展研究，新闻叙事性特征研究以及人才培养等议题越来越受到学界的注重。

结语

本文基于CiteSpace可视化软件，以中国学术期刊网络出版总库（CNKI）收录的数据新闻相关文献为研究对象，总结了2013年以来近十年国内数据新闻研究的发展态势、研究主题和热点变化，得出如下结论。

从2013年开始至2016年，数据新闻领域相关研究处于井喷式的快速发展阶段，这一阶段国内研究学者对数据新闻的研究热度日益高涨。从2017年开始，核心期刊发文量下滑，国内数据新闻研究进入“降温”回落阶段，研究逐渐趋于理性。

国内对数据新闻研究发文量较高的作者多为来自各大高校的新闻传播学界人士，目前已形成以喻国明，方洁，陈昌凤，徐向东等为代表的核心学者。作者间的合作网络均比较松散，尚未形成显著的群体效应。国内数据新闻研究学者更倾向于独立产出成果，积累多元化的创新研究成果。

结合图谱自动聚类结果和文献精读情况，将近十年来国内数据新闻领域研究热点归纳为四大类，分别为数据新闻的基本理念及创新发展研究、数据新闻的生产流程研究、数据新闻的价值及影响研究、数据新闻的人才培养和数据素养研究。

结合关键词时间线视图以及突变词分析，近年来，人工智能、数据科学等新技术在数据新闻中的应用发展研究，数据新闻叙事性特征的回归研究及人才培养等议题越来越受到学界的注重。当前学界的研究热点正在从早期的基本理念、大数据可视化实践等的研究到数据新闻的新技术发展、叙事性回归及人才培养等研究的转变。

参考文献

[1] 陈悦，陈超美，刘则渊等. CiteSpace知识图谱的方法论功能[J]. 科学学研究，2015（2）：242-253.

[2] 任艳莉，王彤. 国内信息素养教育研究知识图谱分析：基于CSSCI论文（2008-2018）[J]. 黑龙江高教研究，2018（9）：36-39.

[3] 方洁，颜冬. 全球视野下的“数据新闻”：理念与实践[J]. 国际新闻界，2013（6）：73-83.

[4] 任瑞娟. 基于数据新闻构建建设性新闻报道框架之可行性研究[J]. 当代传播，2020（4）：45-50.

[5] 刘英华. 开放数据的发展趋势及在数据新闻中的应用[J]. 青年记者，2019（15）：82-83.

[6] 毕秋灵. 数据新闻中的开放数据应用[J]. 湖北社会科学，2016（7）：190-194.

[7] 王琳. 数据新闻与数据价值守护[J]. 中国传媒科技，2022（11）：53-54.

[8] 许向东. 转向、解构与重构：数据新闻可视化叙事研[J]. 国际新闻界，2019（11）：142-155.

[9] 任瑞娟. 预测与发现：数据新闻理论与实践[M]. 北京：科学出版社，2019：78-79.

[10] 史历峰. 国内网媒可视化数据新闻探析[J]. 传媒，2015（20）：48-49.

[11] 许向东. 对中美数据新闻人才培养模式的比较与思考[J]. 国际新闻界，2016（10）：100-109.

[12] 方洁，高璐. 数据新闻：一个亟待确立专业规范的领域——基于国内五个数据新闻栏目的定量研究[J]. 国际新闻界，2015（12）：105-124.

作者简介： 胡岩岩（1986-），女，安徽宿州，初级工程师，研究方向为数据新闻。

（责任编辑：张晓婧）